Joint reconstruction of PET-MRI by exploiting structural similarity

Joint reconstruction of PET-MRI by exploiting structural similarity

Matthias J Ehrhardt[1], Kris Thielemans[2], Luis Pizarro[1], David Atkinson[3], Sébastien Ourselin[1], Brian F Hutton[2] and Simon R Arridge[1]

[1] Centre for Medical Image Computing, University College London, London WC1E 6BT, UK
[2] Institute for Nuclear Medicine, University College London, London NW1 2BU, UK
[3] Centre for Medical Imaging, University College London, London NW1 2PG, UK

E-mail: matthias.ehrhardt.11@ucl.ac.uk

Abstract

Recent advances in technology have enabled the combination of positron emission tomography (PET) with magnetic resonance imaging (MRI). These PET-MRI scanners simultaneously acquire functional PET and anatomical or functional MRI data. As function and anatomy are not independent of one another the images to be reconstructed are likely to have shared structures. We aim to exploit this inherent structural similarity by reconstructing from both modalities in a joint reconstruction framework. The structural similarity between two modalities can be modelled in two different ways: edges are more likely to be at similar positions and/or to have similar orientations. We analyse the diffusion process generated by minimizing priors that encapsulate these different models. It turns out that the class of parallel level set priors always corresponds to anisotropic diffusion which is sometimes forward and sometimes backward diffusion. We perform numerical experiments where we jointly reconstruct from blurred Radon data with Poisson noise (PET) and under-sampled Fourier data with Gaussian noise (MRI). Our results show that both modalities benefit from each other in areas of shared edge information. The joint reconstructions have less artefacts and sharper edges compared to separate reconstructions and the ℓ^2-error can be reduced in all of the considered cases of under-sampling.

(Some figures may appear in colour only in the online journal)

1. Introduction

In recent years the landscape of medical imaging has changed. For many decades single modality scanners to image either anatomy or function have been available and widely used in clinical practice. Anatomy can be imaged for instance by magnetic resonance imaging (MRI) and computer assisted tomography (CT). Often functionality is measured by radioactive labelled markers using positron emission tomography (PET) or single photon emission computed tomography. Based on these single modality scanners, multi-modality scanners were developed to combine the strength of both regimes and image both structure and function. While PET-CT scanners acquire the data for the two modalities sequentially some recently developed PET-MRI scanners are able to perform both scans *simultaneously* [1–3].

This development not only opens up new opportunities for clinicians but also poses a new kind of inverse problem. As 'function follows structure for the most part' [4] the images to be reconstructed are likely to share many edges. Some example images are shown in figure 1 where the reconstructions from PET, MRI and CT of a single patient are shown (using standard settings with the scanner software). The images were affinely registered to a common frame as they were acquired on a sequential PET-CT and on a separate MRI scanner. Despite the fact that all three images show completely different biological information, all three images show *similar structures* due to the same underlying anatomy. Therefore, we propose to *jointly reconstruct* both modalities with the *a priori* knowledge that the images are more likely to show similar structures. Some authors refer to this task as *joint inversion* [5, 6].

We will focus on joint reconstruction of simultaneously acquired PET-MRI data. All methods discussed here are, however, not limited to this application. They take similar structures into account and can therefore be applied to any other multi-modality scanner or even another field of application where similarity in structure is to be expected.

1.1. Contributions

This work is, to the best of our knowledge, the first contribution to *joint* reconstruction in multi-modality medical imaging. We justify the joint objective function used in [5–10] via probabilistic modelling, which also allows arbitrary noise models to be used in a joint reconstruction framework. Moreover, we analyse a general class of joint priors [6–13] and show how they couple the reconstructions in a gradient based optimization scheme. Furthermore, our numerical results show that joint PET-MRI reconstruction can be beneficial to both of the modalities. Most of the results obtained via joint reconstruction are superior, both in visual quality as well as in ℓ^2-error, compared to some widely accepted methods for separate reconstructions.
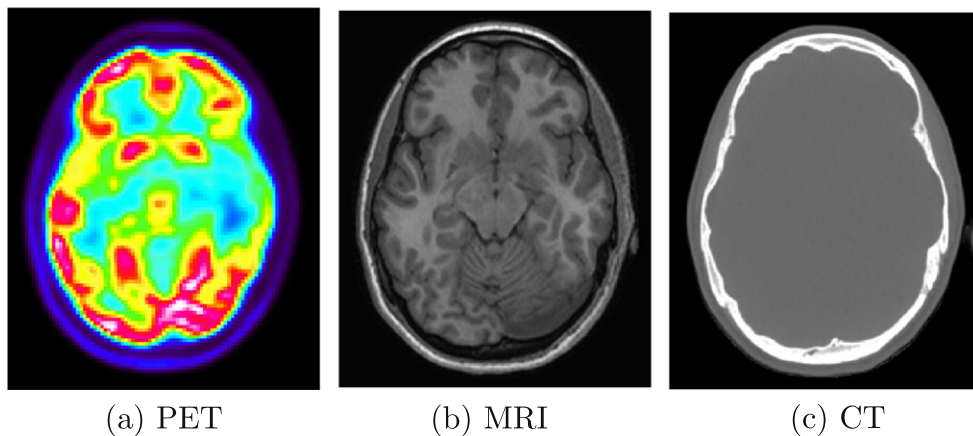
(a) PET (b) MRI (c) CT

Figure 1. Reconstructions of data acquired by PET, MRI and CT of the same patient. Although the images contain completely different biological information common shapes can be observed due to the same anatomy. Images courtesy of UCL hospital.

1.2. Related Work

Joint reconstruction was, as far as we know, first proposed by Haber and Oldenbourg in 1997 [5] where the application of combined PET and MRI reconstruction was already mentioned but the focus was on geophysical applications. Subsequently, different priors coupling two modalities have been proposed [6–9, 14]. Some of these are related to the methods we use in this paper. We will discuss these relations in section 3.

There is a close connection between joint reconstruction and inverse problems in colour image processing as the colour channels can be seen as different modalities which share information about edges [10, 12, 13, 15–20]. All methods used in this work have been proposed first for colour image processing such as denoising and demosaicing [10, 12, 17].

Also closely related is the work where two different MRI data sets of the same subject are enhanced [21] and reconstructed [22].

Similarly, the reconstruction with *a priori* information from another modality has also been studied. This is sometimes referred to as *model fusion* [6]. In the example of PET and MRI one might want to reconstruct PET using the additional structural information of MRI. On the one hand, one may include this information via a structural prior which gives information about edges [11, 23–28]. On the other hand, theoretical information priors such as joint entropy and mutual information have been used to inform about likely tissue distributions [29–32]. None of these priors have been studied for joint reconstruction.

2. Problem setting and notation

This section is dedicated to introduce the problem of joint reconstruction of PET-MRI more formally and, in addition, it serves to state our notations.

2.1. Positron emission tomography

In PET a radioactive labelled and positron emitting tracer is injected into the patient. As the positrons annihilate with electrons of the tissue two photons in (almost) opposite directions

are sent out which are recorded by detectors around the object. One seeks to reconstruct the tracer distribution from recorded events from pairs of detectors [33, 34]. In a general continuous framework this corresponds to the the x-ray transform which coincides with the Radon transform in 2d [35, 36]. This is of course a simplified model as effects such as scatter are ignored.

Let us denote the activity of the tracer by $u: \Omega \subset \mathbb{R}^N \to [0, \infty)$, where $N = 2$ or 3. We want to reconstruct u on a discretized domain which we denote for the sake of simplicity again by $u \in \mathbb{R}^M$. The discrete PET forward operator $A: \mathbb{R}^M \to \mathbb{R}^K$ describes how likely it is that a pair of photons is detected by one of the detector pairs, i.e., $A_{k,m}$ is the probability for an event at voxel m to be detected by the k-th detector pair. More accurate modelling would include the estimation of scatter, randoms and attenuation [34]. The acquired data in PET are commonly assumed to be an instance of a Poisson distributed vector with expectation Au [33, 34, 37].

2.2. Magnetic resonance imaging

MRI exploits the spin of hydrogen nuclei in the human body (or any other material). A magnetic field makes these randomly oriented spins align so that the position of the spin can be encoded by tiny variations in their resonant frequency. The collected signal corresponds to the modulated spin density integrated over the whole domain. The inverse problem in MRI is to reconstruct an image from these Fourier measurements [38, 39].

MRI is a very versatile imaging modality. Depending on the imaging sequence used for acquisition, the MRI image can reflect the proton density of the tissue or some other tissue related parameter [38, 39]. Independent of the imaging sequence we will denote the MRI image by $v: \Omega \to \mathbb{C}$. We denote the discrete MRI forward operator by $B: \mathbb{C}^M \to \mathbb{C}^L \simeq \mathbb{R}^{2L}$ which consists of the Fourier transform followed by a measurement or sampling operator being usually a projection on the measured frequencies.

The noise in MRI is commonly modelled as additive Gaussian with standard deviation depending on, among others, the temperature, the receiver bandwidth and the field strength so that we acquire data $g \sim \mathcal{N}(Bv, \sigma I)$, $\sigma > 0$ [40–42].

It became popular to speed up the MRI data acquisition by omitting 'unnecessary' measurements and 'replacing' them using *a priori* knowledge. This concept known as *compressed sensing* [43–46] has been already applied to MRI with great success [22, 47, 48]. Successfully exploited *a priori* knowledge in MRI includes, for instance, sparsity in the gradient or wavelet domain [47], sparsity in a self-learned dictionary [48] or similarity from different MRI contrasts [22]. In our contribution, the additional *a priori* knowledge is the similarity of MRI to simultaneously acquired PET.

For simplicity we assume the phase in MRI to be zero and therefore restrict v to be real-valued.

2.3. Probabilistic modelling for joint reconstruction

As we have now introduced both modalities PET and MRI we can discuss how they can be linked. We formulate our beliefs using probabilistic modelling and graphical models [49, 50]. The PET and MRI images u, v are random variables. Our intuitive belief about these unknown images u, v are described in the graphical model shown in figure 2. On the one hand, both u and v depend on a common object and are therefore not independent of each other. The data are modelled as random variables with expectation linearly depending on these images. Hence, also the data of PET f and MRI g are not independent of another. On the
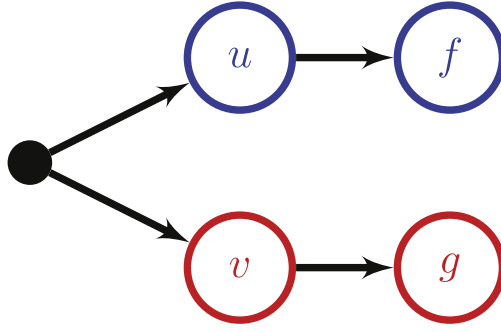
Figure 2. A graphical model encodes our beliefs for joint reconstruction in PET-MRI. See the text for further details.

other hand, if we have knowledge about u, v does not provide extra information about f and vice versa. This means, that f and g are conditionally independent given u, v. Formally, this leads to the separation of the multi-modality likelihood

$$p(f, g \mid u, v) = p(f \mid u, v)p(g \mid u, v) = p(f \mid u)p(g \mid v). \tag{1}$$

The posterior probability of the PET-MRI image (u, v) for observed PET-MRI data (f, g) can then be simplified as

$$p(u, v \mid f, g) \propto p(f \mid u)p(g \mid v)p(u, v) \tag{2}$$

using Bayes' rule. When we further assume white, complex Gaussian noise in MRI, Poisson noise in PET and a prior of the form $p(u, v) = \exp(-\alpha \mathcal{R}(u, v))$ then maximizing (2) is equivalent to minimizing

$$\mathcal{J}(u, v) = \sum_i (Au)_i - f_i \log (Au)_i + \frac{1}{2\sigma^2} |Bv - g|^2 + \alpha \mathcal{R}(u, v), \tag{3}$$

where $|x| := \left(\sum_i |x_i|^2 \right)^{1/2}$ denotes the ℓ^2-norm.

Remark 2.1. The joint objective functional (3) but with Gaussian noise in both modalities coincides with the functional used in [5–10] up to a scaling of the data fidelity terms.

3. Methods for joint reconstruction

This section describes the models for our prior $\mathcal{R}$ which couples the reconstruction of PET and MRI.

3.1. Parallel level sets

3.1.1. Measuring parallelism. A general concept of coupling structure in a multi-channel image was proposed [10] with special cases including [6, 7, 12]. Structure in an image can be identified with its level sets

$$\{x \in \Omega \mid u(x) = s\}. \tag{4}$$

In differentiable images the gradient is perpendicular to the level sets, so similar structures can be measured by similarity of the gradients' orientation. Therefore, parallelism of the gradients at each spatial location is a measure of structural similarity of two images.

3.1.2. Measuring parallelism of vectors. It is well known that for any vectors $x, y \in \mathbb{R}^N$ the following relation is fulfilled

$$|\langle x, y \rangle| = |\cos(\theta)| \, |x| \, |y|, \tag{5}$$

with θ denoting the angle between them in the plane they span and $\langle x, y \rangle := \sum_i x_i y_i$ the Euclidean scalar product. Consequently, one can determine how parallel x and y are by the 'distance measure'

$$d(x, y) := |x| \, |y| - |\langle x, y \rangle|. \tag{6}$$

It is obvious that d is symmetric and $d(x, y) \geqslant 0$ with equality if and only if x is parallel to y. These properties guarantee that d is suitable to measure parallelism. Moreover, it can be generalized to

$$d_{\varphi,\psi}(x, y) := \varphi(\psi(|x| \, |y|) - \psi(|\langle x, y \rangle|)) \tag{7}$$

with arbitrary strictly increasing functions $\varphi, \psi : [0, \infty) \to [0, \infty)$. Obviously, $d_{\varphi,\psi}$ is still symmetric, non-negative and only parallel vectors have minimal 'distance'.

3.1.3. Measuring parallelism of images. This concept of parallel vectors can be easily applied to (differentiable) images. Instead of arbitrary vectors we measure how parallel their gradients are at any point in their domain $x \in \Omega$, i.e.,

$$d_{\varphi,\psi}(\nabla u(x), \nabla v(x)) = \varphi(\psi(|\nabla u(x)| \, |\nabla v(x)|) - \psi(|\langle \nabla u(x), \nabla v(x) \rangle|)). \tag{8}$$

Hence, we end up with a global measure of how 'similar' two images u and v are

$$\mathrm{PLS}(u, v) = \int_\Omega d_{\varphi,\psi}(\nabla u(x), \nabla v(x)) \, \mathrm{d}x. \tag{9}$$

In order to make the notation more easily accessible we will suppress the explicit spatial dependence from now on.

In homogeneous regions of v the gradient ∇v vanishes and this functional does not provide any information for u. Therefore, one instead can take the modified version

$$d_{\varphi,\psi,\beta}(\nabla u, \nabla v) := \varphi\left(\psi\left(|\nabla u|_\beta \, |\nabla v|_\beta \right) - \psi\left(|\langle \nabla u, \nabla v \rangle|_{\beta^2} \right) \right) \tag{10}$$

proposed in [10] with the 'smoothed norm' $|x|_\beta := (|x|^2 + \beta^2)^{1/2}$ for some $\beta > 0$ and the global measure reads

$$\mathrm{PLS}_\beta(u, v) = \int_\Omega d_{\varphi,\psi,\beta}(\nabla u, \nabla v). \tag{11}$$

Remark 3.1. The choice how to smooth the norm is rather arbitrary. Other common choices are the Huber or the log-cosh approximation.

Remark 3.2. In general, one might want to use different smoothing parameters β and γ for $|x|_\beta$ and $|y|_\gamma$ but to keep the notation simple we used the same for both modalities.

Example 3.3. In [10] we applied (11) with $\varphi, \psi = id$ to denoising and demosaicking of colour images. We will refer to this prior as *linear parallel level sets*. Another choice which was used in [10] is defined by $\varphi(s) = (1 + s)^{1/2}, \psi(s) = s^2$. Due to the quadratic nature of ψ we call it *quadratic parallel level sets*. This functional was first introduced in [12, 13] where it was motivated by high energy physics. It was observed in [10] that linear parallel level sets leads to sharper images compared to the quadratic version. We will give a theoretical justification for this observation in section 4.

Example 3.4. Another choice used previously in geophysics is (9) with $\varphi = id$ and $\psi(s) = s^2$ [6]. This functional coincides with the squared magnitude of the cross product of the gradients proposed by [7–9]. Similar functions have been applied to registration [51, 52] and face recognition [53]. We will only briefly discuss this choice further here because it possesses unfavourable noise suppression properties.

Remark 3.5. The functional proposed in [11] can be rewritten as

$$\int_\Omega |\nabla u|^2 - \langle \nabla u, w \rangle^2. \tag{12}$$

It is not exactly a special case of the parallel level sets functional but bears lots of similarities. This functional favours alignment of the gradient of u with the normalized and *a priori* defined vector-field w. It is beyond the scope of this paper to investigate such *asymmetric* priors for joint reconstruction.

3.2. Joint total variation

There are many extensions of total variation regularization to multi-modality images. Of importance here is the coupling of the modalities so that one modality can provide information to the other. In the case of two modalities $u, v: \Omega \subset \mathbb{R}^N \to \mathbb{R}$ the *joint total variation* [6] is defined as

$$\mathrm{JTV}(u, v) = \int_\Omega \left(|\nabla u|^2 + |\nabla v|^2 \right)^{1/2}. \tag{13}$$

This functional was first considered as an extension of total variation for colour images [17]. Another generalization of total variation to colour images was proposed in [54]. While the former locally couples the support of the gradients, the latter only provides a global effect. Joint total variation is closely related to block sparsity in the magnitude of the gradient [55–58]. It is known that usual total variation favours sparsity in the gradient domain which is enforced by sparsity in the magnitude of the gradient. Joint total variation favours sparsity in the magnitude of the 'joint gradient' $(\nabla u, \nabla v)$. Therefore, gradients are more likely to occur at the same position. Another explanation why joint total variation can be useful for joint reconstruction is given in the next section.

We will use here a smoothed version of joint total variation so that we can apply tools from smooth optimization, i.e.,

$$\mathrm{JTV}_\beta(u, v) = \int_\Omega \left(|\nabla u|^2 + |\nabla v|^2 + \beta^2 \right)^{1/2}, \tag{14}$$

for some $\beta > 0$. In the notation of the 'smoothed norm' introduced above we can write smooth joint total variation as

$$\mathrm{JTV}_\beta(w) = \int_\Omega |\nabla w|_\beta, \tag{15}$$

where the vector-valued image $w: \Omega \to \mathbb{R}^2$ is defined as $w(x) = (u(x), v(x))$ and the gradient is the Jacobian.

4. Analysis of the generated diffusion process

We have introduced the concepts of parallel level sets and joint total variation to couple the reconstruction of multi-modality images. Priors based on gradient information often lead to a derivative of *diffusion* type [10], i.e., if a prior is of the form $\mathcal{R}(u) = \int \Psi(\nabla u)$ with $\nabla \Psi(x) = K(x)x$ then the Gâteaux derivative can be identified with

$$D\mathcal{R}[u] = -\mathrm{div}(K(\nabla u)\nabla u). \tag{16}$$

The generated diffusion process depends crucially on the *diffusivity* $K(\nabla u)$. Its analysis can give more insight how a prior influences the final reconstruction when an algorithm based on derivative information is used.

We will see in the following that in both cases the Gâteaux derivative leads to a non-linear, inhomogeneous diffusion, i.e., the diffusivity depends on the image and the location. Moreover, in case of parallel level sets the diffusion will be anisotropic, i.e., it depends on the *direction* [59, 60].

4.1. Parallel level sets

In order to analyse the diffusion generated by parallel level sets we first need to compute the Gâteaux derivative. Most of the derivations have already been shown in [10].

Let us denote the space of continuously differentiable functions with zero Neumann boundary condition by $C_*^1(\Omega)$, i.e., the derivative with respect to the normal of the boundary being zero.

Lemma 4.1. *([10]). Let φ, ψ be continuously differentiable functions. The Gâteaux derivative of* $\mathrm{PLS}_\beta : C_*^1(\Omega) \times C_*^1(\Omega) \to \mathbb{R}$ *defined by (11) at (u, v) is given by*

$$D\mathrm{PLS}_\beta(u, v) = -\mathrm{div}\left[\begin{pmatrix} \kappa(\nabla u, \nabla v) & \tau(\nabla u, \nabla v) \\ \tau(\nabla u, \nabla v) & \kappa(\nabla v, \nabla u) \end{pmatrix}\begin{pmatrix} \nabla u \\ \nabla v \end{pmatrix}\right] \tag{17}$$

with diffusivities

$$\kappa(x, y) := \rho(x, y)|y|_\beta |x|_\beta^{-1} \psi'\left(|x|_\beta |y|_\beta\right) \geqslant 0, \tag{18}$$

$$\tau(x, y) := -\rho(x, y)\langle x, y\rangle |\langle x, y\rangle|_{\beta^2}^{-1} \psi'\left(|\langle x, y\rangle|_{\beta^2}\right) \in \mathbb{R}, \tag{19}$$

$$\rho(x, y) := \varphi'\left(\psi\left(|x|_\beta |y|_\beta\right) - \psi\left(|\langle x, y\rangle|_{\beta^2}\right)\right) \geqslant 0. \tag{20}$$

Remark 4.2. The derivatives and the matrix operations have to be applied component-wise so that the notation in (17) makes sense.

The flow generated by (17) is difficult to analyse because of the cross-diffusivities τ. Therefore, we write an equivalent formulation which does not rely on cross-diffusivities but on matrix-valued diffusivities.

Proposition 4.3. *With the same assumptions as in lemma 4.1 the Gâteaux derivative can be equivalently rewritten as*

$$DPLS_\beta(u, v) = -\mathrm{div}\begin{pmatrix} K(\nabla u, \ \nabla v)\,\nabla u \\ K(\nabla v, \ \nabla u)\,\nabla v \end{pmatrix}, \tag{21}$$

where the diffusivities are given by

$$K(x, y) := \kappa(x, y)I + \mu(x, y)yy^T \in \mathbb{R}^{N\times N}, \tag{22}$$

$$\mu(x, y) := -\rho(x, y)|\langle x, y\rangle|_{\beta^2}^{-1}\,\psi'\left(|\langle x, y\rangle|_{\beta^2}\right) \leqslant 0. \tag{23}$$

and κ and ρ by (18) and (20), respectively.

Proof. The assertion for the first equation in (21) follows immediately from the fact that $\langle \nabla u, \ \nabla v\rangle\,\nabla v = (\nabla v\,\nabla v^T)\,\nabla u$. Interchanging the roles of u and v completes the proof. $\qquad\square$

Remark 4.4. The lemma 4.1 was proven only for PLS_β. Under similar conditions on φ and ψ it also holds for PLS with $|\cdot|_\beta$ being replaced by $|\cdot|$. Hence, this proposition can be applied to PLS as well.

For the analysis of the diffusivity we need to remind the reader about the orthogonal complement and the span.

Definition 4.5. For any $y \in \mathbb{R}^N$ we can decompose the space as $\mathbb{R}^N = y^\| \oplus y^\perp$ where *span* and the *orthogonal complement* of y are defined as

$$y^\| := \left\{ sy \in \mathbb{R}^N \colon s \in \mathbb{R} \right\} \tag{24}$$

and

$$y^\perp := \left\{ x \in \mathbb{R}^N \colon \langle x, y\rangle = 0 \right\}. \tag{25}$$

Lemma 4.6. *For all matrices $K = \kappa I + \mu yy^T \in \mathbb{R}^{N\times N}$ with $\kappa, \mu \in \mathbb{R}$ the following holds.*

 (i) *K is symmetric.*
(ii) *All $x \in y^\|$ are eigenvectors to the eigenvalue $\lambda^\| := \kappa + \mu |y|^2$.*
(iii) *All $x \in y^\perp$ are eigenvectors to the eigenvalue $\lambda^\perp := \kappa$.*

Proof. Ad (i) There is

$$K^T = \left(\kappa I + \mu yy^T\right)^T = \kappa I^T + \mu\left(yy^T\right)^T = \kappa I + \mu yy^T = K. \tag{26}$$

Ad (ii) Let $x \in y^{\parallel}$. So there is $x = sy$ for some $s \in \mathbb{R}$ and hence

$$Kx = sKy = s\left(\kappa I + \mu yy^T\right)y = s\left(\kappa y + \mu |y|^2 y\right) \tag{27}$$

$$= \left(\kappa + \mu |y|^2\right)x = \lambda^{\parallel}x. \tag{28}$$

Ad (iii) Let $x \in y^{\perp}$. Then there is

$$Kx = \left(\kappa I + \mu yy^T\right)x = \kappa x + \mu y \langle y, x \rangle = \kappa x = \lambda^{\perp}x. \tag{29}$$

$\square$

To analyse the diffusion we need to introduce local coordinates for the image domain.

Definition 4.7. (local coordinates). Let $f\colon \Omega \subset \mathbb{R}^N \to \mathbb{R}$ be a differentiable function. We denote by $R[f]$ a local coordinate mapping $R[f]\colon \Omega \to \mathbb{R}^{N \times N}$ such that for any point $x \in \Omega$ the columns of $R[f](x)$ form a basis of $\nabla f(x)^{\parallel} \oplus \nabla f(x)^{\perp}$.

Theorem 4.8. *Let $R[u]$ and $R[v]$ be local coordinates for u and v. Then the diffusion (21) can be rewritten as*

$$\mathrm{DPLS}_\beta[u, v] = -\mathrm{div}\begin{pmatrix} R[v]\,\Lambda[u, v]\,R[v]^T\,\nabla u \\ R[u]\,\Lambda[v, u]\,R[u]^T\,\nabla v \end{pmatrix}, \tag{30}$$

where the diffusivities are defined as

$$\Lambda[u, v] := \mathrm{diag}\left(\lambda^{\parallel}(\nabla u, \nabla v),\ \lambda^{\perp}(\nabla u, \nabla v),\ \ldots,\ \lambda^{\perp}(\nabla u, \nabla v)\right) \tag{31}$$

with

$$\lambda^{\parallel}(x, y) := \kappa(x, y) + \mu(x, y)|y|^2 \in \mathbb{R} \qquad \lambda^{\perp}(x, y) := \kappa(x, y) \geqslant 0 \tag{32}$$

for κ and μ as defined in (18) and (23).

Proof. This result follows immediately from proposition 4.3 combined with lemma 4.6. $\square$

Remark 4.9. Theorem 4.8 tells us that the diffusion for the image u generated by the parallel level sets functional has principal directions given by the other image v and vice versa. Moreover, the directions but not the amount of the diffusion are independent of the functions φ and ψ.

Remark 4.10. In terms of the function $\tilde{\psi}(s) := \psi'(s)/s$ the diffusivity in the direction of the image gradient can be rewritten as

$$\lambda^{\parallel}(x, y) = \rho(x, y)\left(|y|_\beta^2\,\tilde{\psi}\left(|x|_\beta\,|y|_\beta\right) - |y|^2\,\tilde{\psi}\left(|\langle x, y \rangle|_{\beta^2}\right)\right) \tag{33}$$

so the sign of the diffusivity depends on the monotonicity of $\tilde{\psi}$. This explains our choice to name the different methods of parallel level sets by the choice of ψ.

Example 4.11. (Linear Parallel Level Sets). The principal diffusivities for the linear parallel level sets functional, i.e., (11) with $\varphi, \psi = id$, are given by

$$\lambda^{\parallel} = |y|_\beta \, |x|_\beta^{-1} - |y|^2 \, |\langle x, y \rangle|_{\beta^2}^{-1} \tag{34}$$

$$= -|y|_\beta \, |x|_\beta^{-1} \, |\langle x, y \rangle|_{\beta^2}^{-1} \left(\frac{|x|_\beta \, |y|}{|x| \, |y|_\beta} \, |x| \, |y| - \frac{|\langle x, y \rangle|_{\beta^2}}{|\langle x, y \rangle|} \, |\langle x, y \rangle| \right) \tag{35}$$

$$\lambda^{\perp} = |y|_\beta \, |x|_\beta^{-1} > 0. \tag{36}$$

It is of interest to analyse the four cases

(i) vanishing gradient information, i.e., $|x| \to 0$,

(ii) vanishing side information, i.e., $|y| \to 0$,

(iii) large gradients, i.e., $\beta \to 0$, and

(iv) large and aligned gradients, i.e., $\beta \to 0$, $|\langle x, y \rangle| = |x| \, |y|$,

where we always assume that $|x|, |y|, |\langle x, y \rangle|, \beta > 0$. It is easy to compute that the principal diffusivities are then

(i) $\qquad \lambda^{\parallel} \to \beta^{-1} \left(|y|_\beta - \beta^{-1} |y|^2 \right) \qquad\qquad \lambda^{\perp} \to \beta^{-1} |y|_\beta$

(ii) $\qquad \lambda^{\parallel} \to \beta \, |x|_\beta^{-1} > 0 \qquad\qquad\qquad \lambda^{\perp} \to \beta \, |x|_\beta^{-1}$

(iii) $\lambda^{\parallel} \to - |y| \, |x|^{-1} \, |\langle x, y \rangle|^{-1} (|x| \, |y| - |\langle x, y \rangle|) \leqslant 0 \; \lambda^{\perp} \to |y| \, |x|^{-1}$

(iv) $\qquad\qquad \lambda^{\parallel} \to 0 \qquad\qquad\qquad\qquad \lambda^{\perp} \to |y| \, |x|^{-1}.$

Example 4.12. (Quadratic Parallel Level Sets). The principal diffusivities for the quadratic parallel level sets functional, i.e., (11) with $\varphi(s) = \sqrt{s+1}$ and $\psi(s) = s^2$ are given by

$$\lambda^{\parallel} = \beta^2 \left(1 + |x|_\beta^2 \, |y|_\beta^2 - |\langle x, y \rangle|_{\beta^2}^2 \right)^{-1/2} > 0 \tag{37}$$

$$\lambda^{\perp} = |y|_\beta^2 \left(1 + |x|_\beta^2 \, |y|_\beta^2 - |\langle x, y \rangle|_{\beta^2}^2 \right)^{-1/2} > 0. \tag{38}$$

For the special cases considered in example 4.11 we get

(i) $\lambda^{\parallel} \to \beta^2 \left(1 + \beta^2 |y|^2 \right)^{-1/2} \qquad \lambda^{\perp} \to |y|_\beta^2 \left(1 + \beta^2 |y|^2 \right)^{-1/2}$

(ii) $\lambda^{\parallel} \to \beta^2 \left(1 + \beta^2 |x|^2 \right)^{-1/2} \qquad \lambda^{\perp} \to \beta^2 \left(1 + \beta^2 |x|^2 \right)^{-1/2}$

(iii) $\qquad \lambda^{\parallel} \to 0 \qquad\qquad \lambda^{\perp} \to |y|^2 \left(1 + |x|^2 \, |y|^2 - \langle x, y \rangle^2 \right)^{-1/2}$

(iv) $\qquad \lambda^{\parallel} \to 0 \qquad\qquad\qquad \lambda^{\perp} \to |y|^2.$

Remark 4.13. Example 4.11 shows that linear parallel level sets yields in some cases negative principal diffusivities. This is commonly referred to as *backward diffusion*. Backward diffusion can lead to instabilities but also might result in sharper images. The principal diffusivities of quadratic parallel level sets are all positive which makes it more stable. This analysis explains the observation in [10] where linear parallel sets resulted in sharper images compared to quadratic parallel level sets.

Remark 4.14. These examples also show that the diffusion is truely anisotropic as the diffusion along and perpendicular to the gradient are different except for the case when the side information vanishes.

Example 4.15. The principal diffusivities for example 3.4 are given by

$$\lambda^{\parallel} = 0 \tag{39}$$

$$\lambda^{\perp} = |y|^2. \tag{40}$$

The flow generated is very simple as there is only diffusion only along the edges. It stops when side information is flat and is therefore not able to remove noise in this case.

4.2. Joint total variation

In this section we discuss the diffusion which is generated by joint total variation. It follows immediately from the well known Gâteaux derivative of total variation (see [61]) that the Gâteaux derivative of joint total variation can be identified with

$$DJTV_\beta[u, v] = -\text{div}\begin{pmatrix} \eta(u, v)\, \nabla u \\ \eta(u, v)\, \nabla v \end{pmatrix}. \tag{41}$$

where the diffusivity for both modalities is

$$\eta(u, v) := \left(|\nabla u|^2 + |\nabla v|^2 + \beta^2 \right)^{-1/2}. \tag{42}$$

The diffusivity for both modalities are the same. It reduces to zero when either ∇u or ∇v is large, hence information about an edge can be transferred from one modality to the other.

The diffusion of joint total variation is isotropic as the diffusivities are scalar-valued. Similarly to usual total variation [62] one can consider joint total variation with anisotropic diffusion but this is out of the scope of this paper.

5. Numerical experiments

5.1. Technical details

5.1.1. Implementation. In our experiments we aim to compute minimizers of (3) with and without an explicit prior. If no explicit prior is used we terminate the iterations before convergence to regularize the reconstruction. The priors we used include total variation, joint total variation (see section 3.2) and linear and quadratic parallel level sets (see section 3.1). As we want to compare the different priors for joint reconstruction we chose to minimize the functional with the same quasi-Newton method (L-BFGS-B [63], software available at [64] with mex wrapper for use in MATLAB [65]) in every case. The MATLAB implementation is available as supplementary data from stacks.iop.org/ip/31/015001/mmedia can be obtained from the authors' homepage [66].

Quasi-Newton methods rely on gradient information only and approximate the Hessian. The gradient of total variation, joint total variation (41), quadratic and linear parallel level (21) sets all are of diffusion type. Hence, we need to discretize gradients and divergences. Gradients are discretized with forward differences and the divergence with backward differences so that the discretized gradient corresponds to the gradient of the discretized functional.
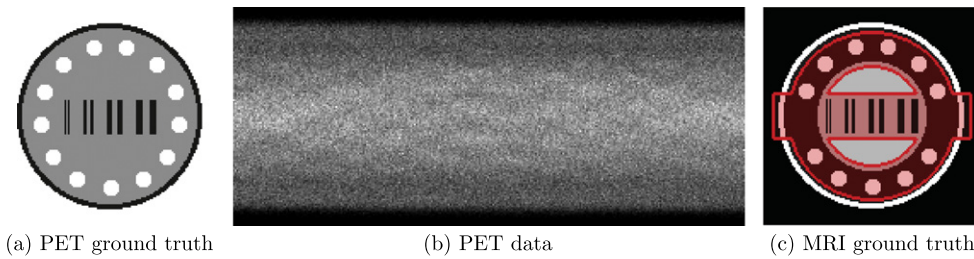
(a) PET ground truth   (b) PET data   (c) MRI ground truth

Figure 3. The figure shows the ground truth of PET and MRI as well as the PET data. The region of interest for the quantatitive analysis is marked in red.

5.1.2. Data sets. We consider the setting of joint PET and MRI reconstruction. The PET operator is modelled to be the discrete x-ray transform taken from [67]. To model the lack of resolution the operator consists also of a convolution in the sinogram domain in the radial direction with a Gaussian of full width half maximum of five pixels (for sinograms of size 128×300). The ground truth models for both modalities together with the blurry sinogram for PET are shown in figure 3. Six different cases of MRI undersampling were tested: the first case of sampling is a fully sampled MRI (`full`). For the second and third sampling the data were acquired along 20 and 15 radial lines in k-space (`radial20`, `radial15`). The fourth and fifth case are spirals with either uniform density (`spiralUni`) or a higher density in the high frequencies (`spiralHigh`). In the last test case we performed Cartesian sampling but only acquired every second line (`lines2`). The MRI data are presented next to the results. Gaussian noise with expected energy of 4% is added to the MRI. The PET data is an instance of a multi-variate Poisson distribution with an expected number of `1e+6` counts in total.

5.1.3. Parameter selection. As both the linear and the quadratic parallel level set priors are non-convex (see appendix A) we might end up with a local minimizer. Therefore, a good initial guess is important to end up at a desirable local minimizer.

The parameters α and β are chosen to minimize the relative ℓ^2-error of the reconstruction x and the ground truth x^*

$$e(x) := |x - x^*|/|x^*|, \tag{43}$$

where the norms are only taken over the region of interest, see figure 3. This region of interest was chosen to give no weight to the background and less to large uniform regions as well as the outer boundary.

5.2. Results

The joint reconstruction approaches described in section 3 are compared against separated reconstruction which is performed either by total variation or having no explicit prior. The latter means early stopping for PET or minimal energy (zero filled) reconstruction for MRI.

5.2.1. Visual assessment. Figure 4 shows the results for the case `full` with the reconstructed images shown on the left and the error images (compared to the ground truth) on the right where blue indicates over and red underestimation. The MRI reconstructions are visually perfect for all priors as it can be seen from both the images themself and the error images. The results for PET differ a lot between the different priors. When the iterations have been terminated the reconstruction looks the worst. Most of the introduced artefacts are
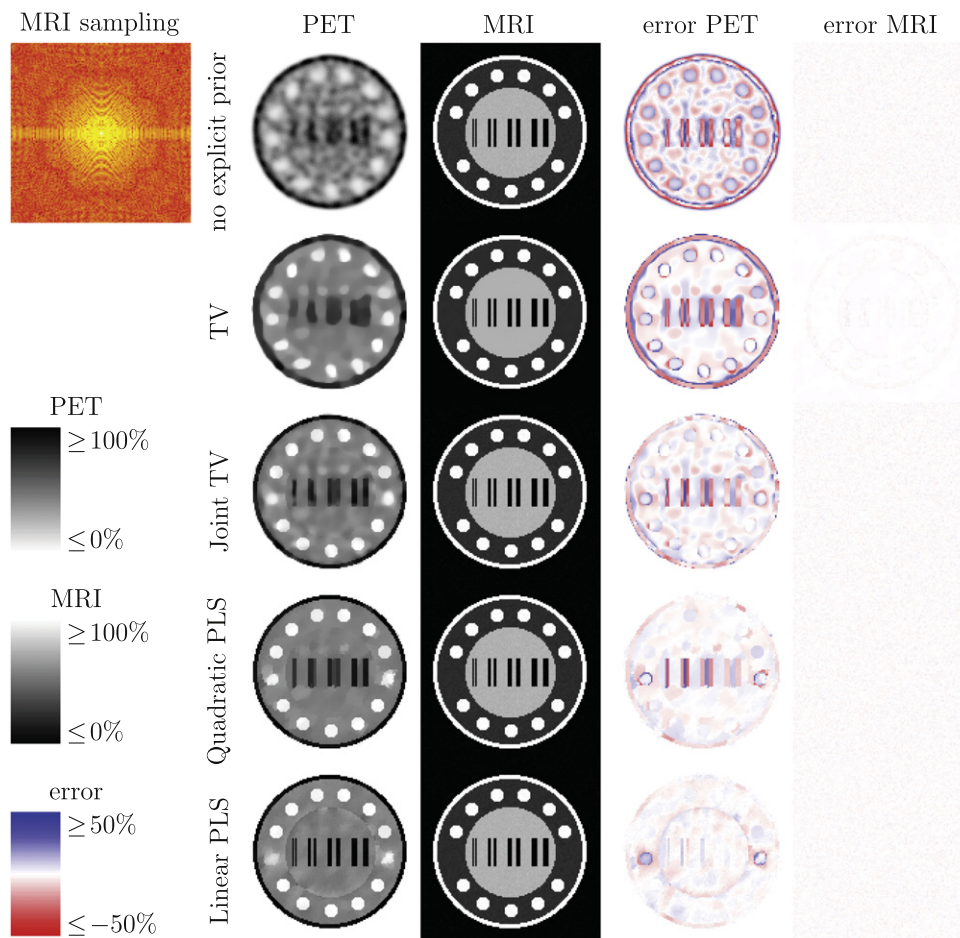
Figure 4. Reconstruction results for the case `full`, i.e. MRI is fully sampled. The reconstructed images are shown in gray scales on the left and the error images in red and blue on the right where blue indicates over and red under estimation. The methods' abbreviations are TV=total variation, PLS=parallel level sets.

removed when using total variation but the resolution in PET remains more or less the same. Coupling PET and MRI with joint total variation results in a superior PET image compared to total variation. Shared edges are better reconstructed using quadratic parallel level sets. In the case of linear parallel level sets almost no errors at shared edges can be observed. In areas of shared information the images look almost indistinguishable from the ground truth. The edges which are not shared between the two modalities introduce a fake edge artefact. From the error images in PET we see that the errors are decreasing from top to bottom. The errors for linear parallel level sets are almost only in regions of non-shared edge information; both coincide with the visual impression on the left.

The results for `radial20` are shown in figure 5. With only 20 radial spokes a zero-filled reconstruction does not give a reasonable reconstruction for MRI any more but including total variation into the reconstruction resolves most of it. While joint total variation and quadratic parallel level sets are not able to improve on this linear parallel level sets is able to reconstruct
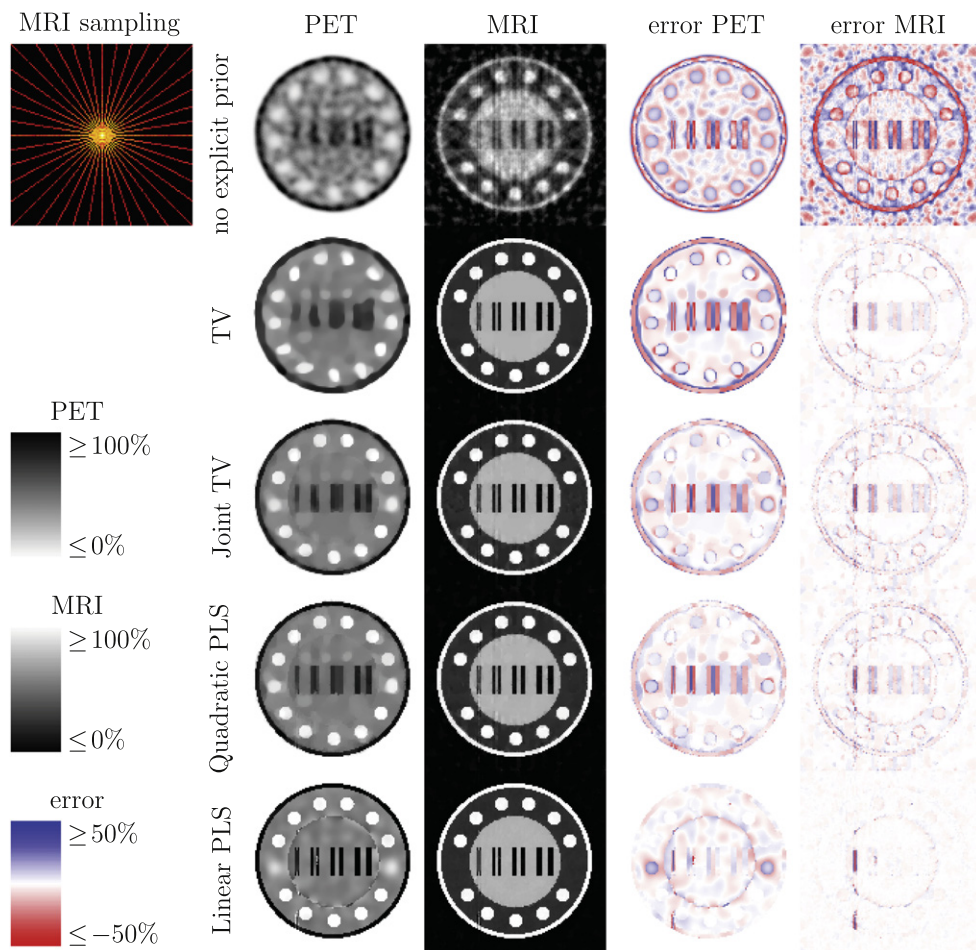
Figure 5. Reconstructions for `radial20` in gray with the error images shown in blue and red on the right. The methods' abbreviations are TV=total variation, PLS=parallel level sets.

with a lot less errors and only the finest lines can not be resolved. The PET reconstructions are similar to the fully sampled case but linear parallel level sets does show a few errors in regions of shared edge information. It can also be seen that the two blobs in PET not being present in MRI are blurrier than total variation and similar to early stopping. This is probably due to the parameter β being constant over and optimized for the whole phantom.

Reducing the number of spokes to 15 (see figure 6) the difference between separate and joint reconstruction becomes even larger with linear parallel level sets still being able to reconstruct reasonable images with only a few errors.

More results are shown in figures 7 and 8 where the MRI data are sampled along spiral trajectories.

The reconstructions from Cartesian undersampling `lines2` are shown in figure 9. While the separated MRI reconstruction shows a two-fold ghosting of the object, joint reconstruction, especially linear parallel level sets, can help to complete the missing
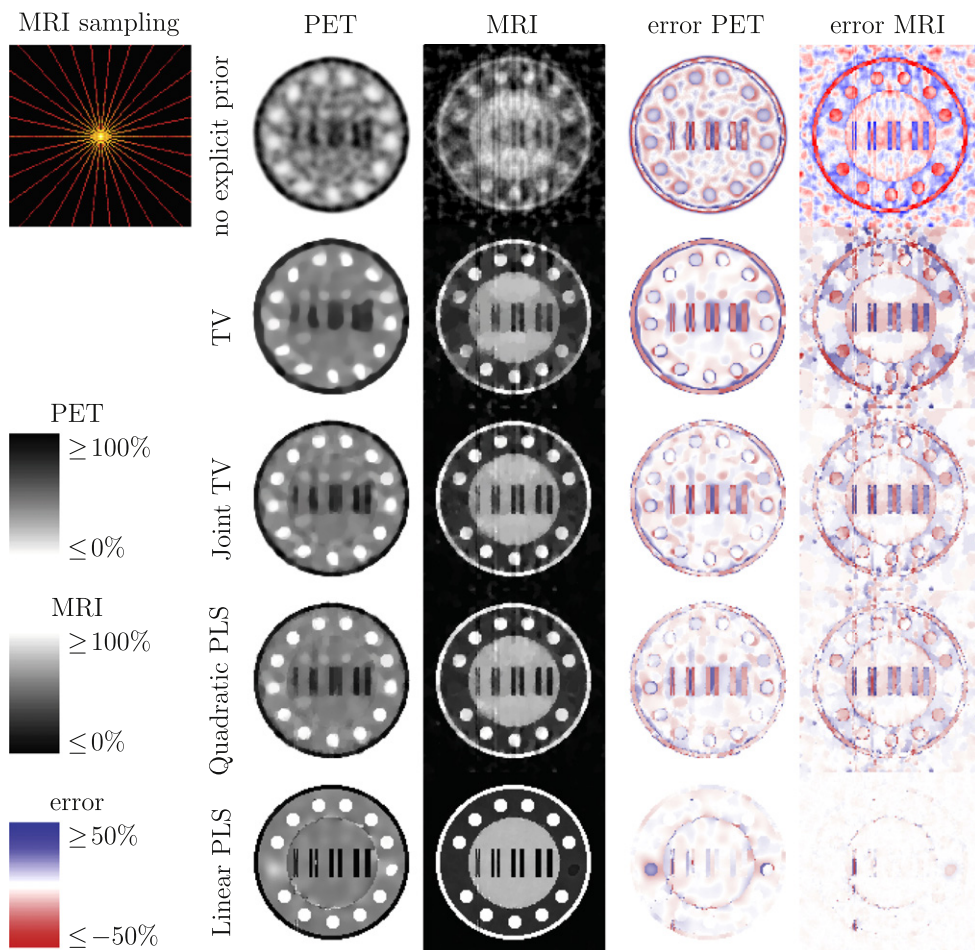
Figure 6. Reconstructions for `radial15` in gray with the error images shown in blue and red on the right. The methods' abbreviations are TV=total variation, PLS = parallel level sets.

information. The ghosting of the inner circle is not fully removed as there is no information in the PET data about this object.

5.2.2. Quantitative assessment. Quantitative results using the relative ℓ^2-error over the *whole* image can be found in figure 10. The errors for for PET are shown on the left and for MRI on the right. For PET it is easy to see that for all cases of sampling separate reconstruction is always worse than joint reconstruction with joint total variation being weaker than parallel level sets. Among the two parallel level sets priors the linear one performs much better. Furthermore, it can be seen that the results are quite robust on the kind and amount of under-sampling.

For full sampling in MRI all methods give very good results. When MRI is reconstructed from under-sampled data using the zero-filled Fourier transform the results are always the worst. Total variation performs as well as joint total variation and quadratic parallel level sets
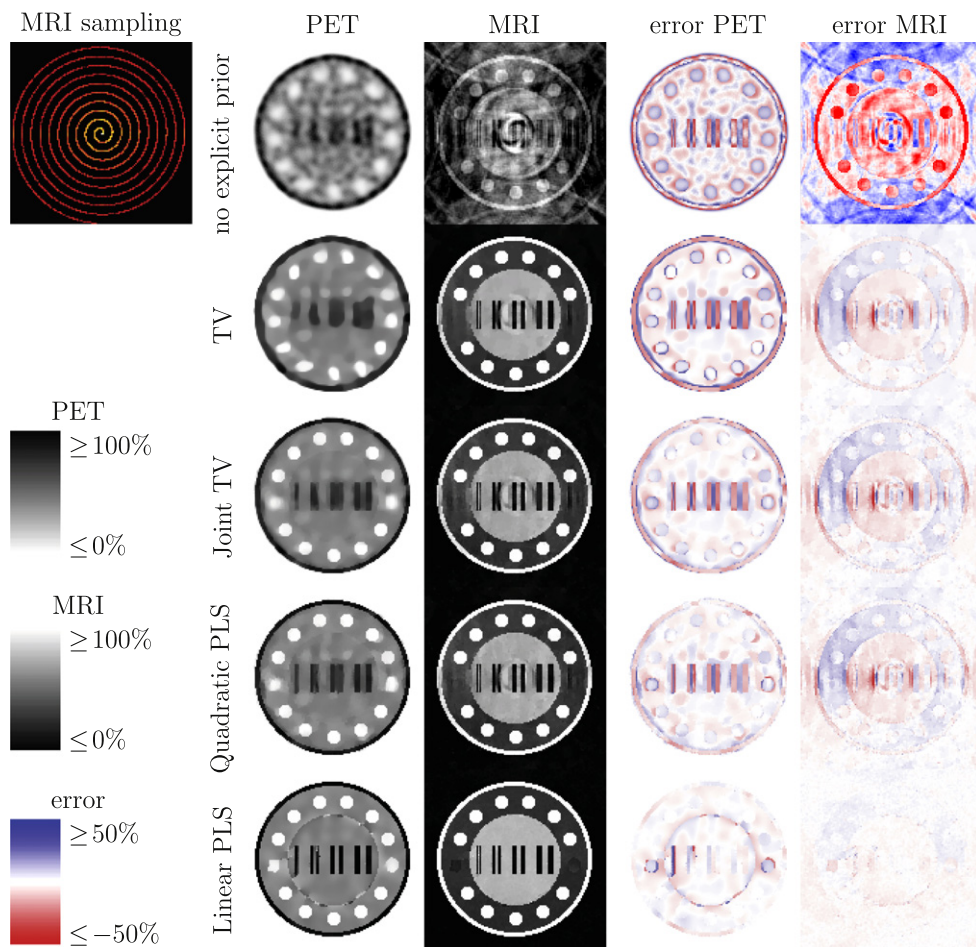
Figure 7. Reconstructions for `spiralUni` on the left and their error images on the right. The methods' abbreviations are TV = total variation, PLS = parallel level sets.

in some cases. The errors of the linear parallel level sets reconstructions are smaller in all the five cases of under-sampling.

6. Conclusion

This work is the first contribution to joint reconstruction in multi-modality medical imaging. We have shown that the previously used framework for joint reconstruction coincides with our belief about multi-modality imaging systems like PET-MRI. Three different priors have been proposed for joint reconstruction of PET-MRI and the corresponding diffusion has been analyzed. This analysis gives more insight on what images are favoured by these priors.

We presented results where blurry Radon data with Poisson noise (PET) and under-determined Fourier data with Gaussian noise (MRI) have been combined in a joint reconstruction process. The numerical results show that our joint reconstruction framework can successfully couple information from these two imaging systems so that both modalities benefit.
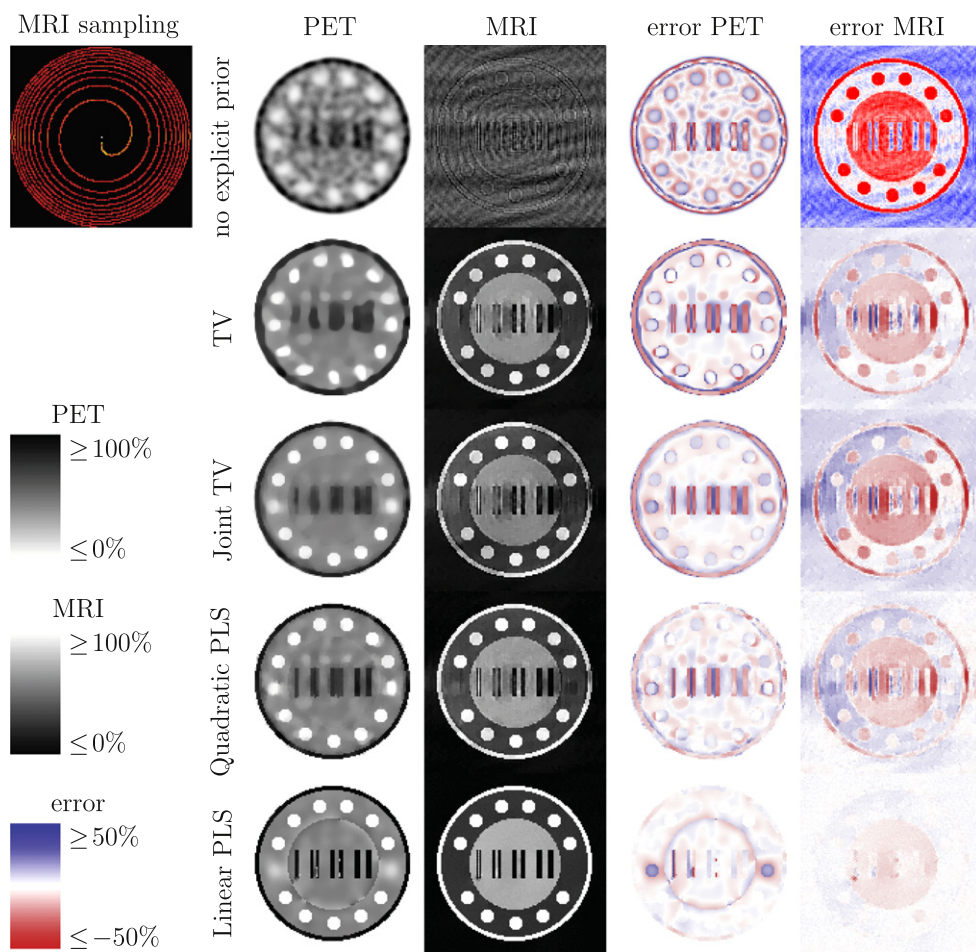
Figure 8. Reconstructions for `spiralHigh` on the left and their error images on the right. The methods' abbreviations are TV = total variation, PLS = parallel level sets.

Similar to multiple priors for one modality [68–70] one might be able to include better the given *a priori* information by using multiple priors. Moreover, the optimization methods have to be adapted to the non-convexity of the parallel level set prior. The effect of noise and non-local parameter tuning for joint reconstruction will also be subject of further research.
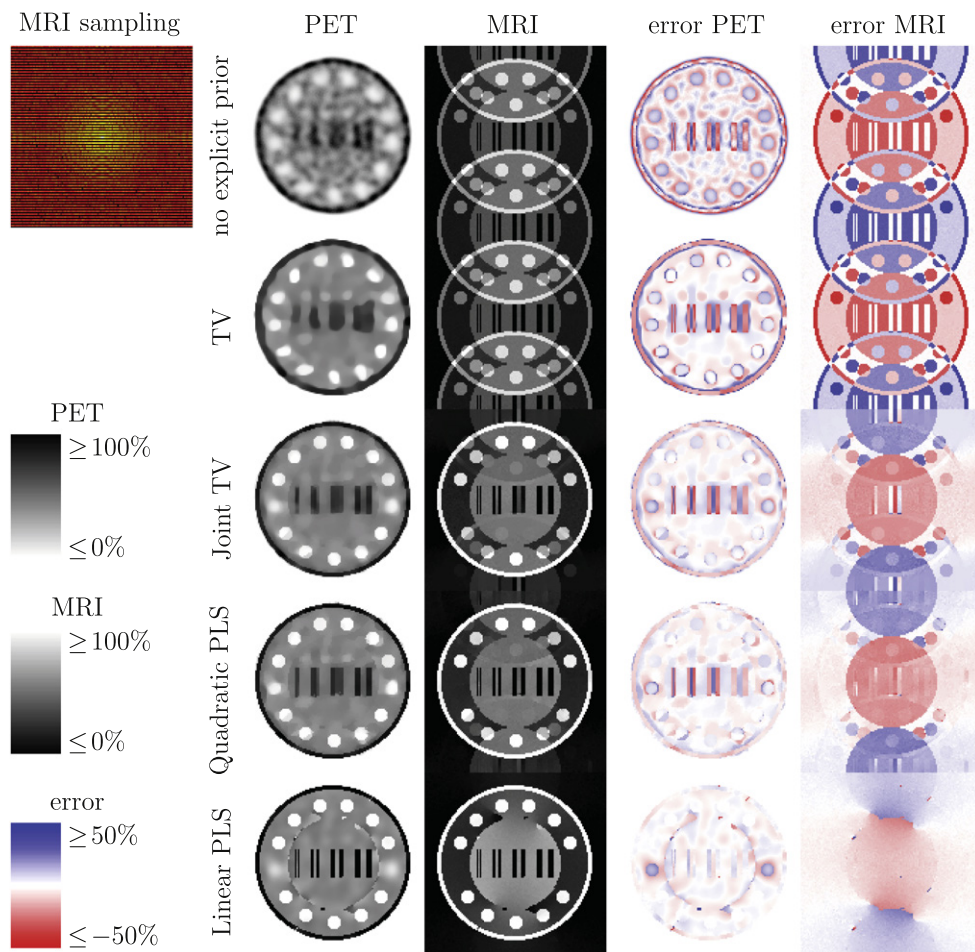
Acknowledgments

Figure 9. The reconstructions for `lines2` are shown in gray on the left and the corresponding error images in blue and red on the right. The methods' abbreviations are TV=total variation, PLS=parallel level sets.

Appendix A. Non-convexity of parallel level sets

In this section we show that all considered versions of the parallel level sets framework are not convex. Under mild conditions a functional $\mathcal{R}(u, v) := \int f(\nabla u, \nabla v)$ is convex if and only if f is convex. We will now analyze f for the special cases considered above.

Proposition A.1. *Let X be a vector space and $f: X \times X \to \mathbb{R}$ a convex functional. Then the growth condition*

$$2f(x/2, y/2) \leqslant f(x, 0) + f(0, y) \tag{A.1}$$
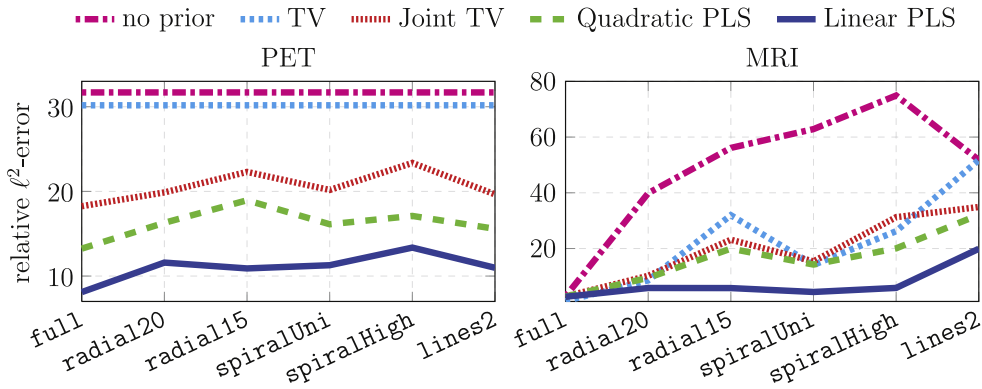
holds for all $x, y \in X$.

Figure 10. The figure shows the relative ℓ^2-errors over the whole phantom (in percent). The results for PET depends only on the sampling in MRI in the case of joint reconstruction. The methods' abbreviations are TV=total variation, PLS=parallel level sets.

Proof. Let $s = 1/2$, $z_0 = (x, 0)$ and $z_1 = (0, y)$. As f is convex there is

$$f(x/2, y/2) = f\left(sz_0 + (1-s)z_1\right) \tag{A.2}$$

$$\leqslant sf(z_0) + (1-s)f(z_1) = 1/2f(x, 0) + 1/2f(0, y) \tag{A.3}$$

and the assertion follows directly. $\qquad\square$

Example A.2. The quadratic parallel level sets functional (see example 3.3) which corresponds to

$$f(x, y) = \left(1 + |x|_\beta^2 |y|_\beta^2 - |\langle x, y\rangle|_{\beta^2}^2\right)^{1/2} \tag{A.4}$$

is not convex in the joint argument.

Proof. Let $x, y \in \mathbb{R}^N$ so that $\langle x, y\rangle = 0$ and $|x| = |y| = 1$. Furthermore, set $\hat{x} := sx$ and $\hat{y} := sy$ for any $s \in \mathbb{R}$. Then the necessary growth condition (A.1) for $f(\hat{x}, \hat{y})$ reads

$$2\left(1 + \beta^2 s^2/2 + s^4/16\right)^{1/2} \leqslant 2\left(1 + \beta^2 s^2\right)^{1/2} \tag{A.5}$$

which does not hold for $s > \sqrt{8}\beta$. $\qquad\square$

Example A.3. The linear parallel level sets functional (see example 3.3) defined via

$$f(x, y) = |x|_\beta |y|_\beta - |\langle x, y\rangle|_{\beta^2} \tag{A.6}$$

is not convex in the joint argument.

Proof. For $x, y \in \mathbb{R}^N$ defined as in the proof of example A.2 the necessary growth condition (A.1)

$$2\left(\beta^4 + \beta^2 s^2/2 + s^4/16\right)^{1/2} - 2\beta^2 \leqslant 2\beta\left(s^2 + \beta^2\right)^{1/2} - 2\beta^2 \tag{A.7}$$

which is again violated for $s > \sqrt{8}\beta$. $\qquad\square$

References

[1] Cherry S R 2006 Multimodality *in vivo* imaging systems: twice the power or double the trouble? *Annu. Rev. Biomed. Eng.* **8** 35–62
[2] Townsend D W 2008 Multimodality imaging of structure and function *Phys. Med. Biol.* **53** 1–39
[3] Catana C, Guimaraes A R and Rosen B R 2013 PET and MR imaging: the odd couple or a match made in heaven? *J. Nucl. Med.* **54** 815–24
[4] Rangarajan A, Hsiao I and Gindi G 2000 A bayesian joint mixture framework for the integration of anatomical information in functional image reconstruction *J. Math. Imaging Vis.* **12** 199–217
[5] Haber E and Oldenburg D W 1997 Joint inversion: a structural approach *Inverse Problems* **13** 63–77
[6] Haber E and Holtzman-Gazit M 2013 Model fusion and joint inversion *Surv. Geophys.* 675–95
[7] Gallardo L A and Meju M A 2003 Characterization of heterogeneous near-surface materials by joint 2d inversion of dc resistivity and seismic data *Geophys. Res. Lett.* **30** 1658
[8] Gallardo L A and Meju M A 2004 Joint two-dimensional dc resistivity and seismic travel time inversion with cross-gradients constraints *J. Geophys. Res.* **109** 1–11
[9] Gallardo L A and Meju M A 2011 Structure-coupled multiphysics imaging in geophysical sciences *Rev. Geophys.* **49** 1–19
[10] Ehrhardt M J and Arridge S R 2014 Vector-valued image processing by parallel level sets *IEEE Trans. Image Process.* **23** 9–18
[11] Kaipio J P, Kolehmainen V, Vauhkonen M and Somersalo E 1999 Inverse problems with structural prior information *Inverse Problems* **15** 713–29
[12] Sochen N, Kimmel R and Malladi R 1998 A general framework for low level vision *IEEE Trans. Image Process.* **7** 310–8
[13] Kimmel R, Malladi R and Sochen N 2000 Images as embedded maps and minimal surfaces: movies, color, texture, and volumetric medical images *Int. J. Comput. Vis.* **39** 111–29
[14] Aghasi A, Mendoza-Sanchez I, Miller E L, Ramsburg C A and Abriola L M 2013 A geometric approach to joint inversion with applications to contaminant source zone characterization *Inverse Problems* **29** 115014
[15] di Zenzo S 1986 A note on the gradient of a multi-image *Comput. Vis. Graph. Image Process.* **33** 116–25
[16] Whitaker R T 1993 Geometry-limited diffusion in the characterization of geometric patches in images *CVGIP, Image Underst.* **57** 111–20
[17] Sapiro G and Ringach D L 1996 Anisotropic diffusion of multivalued images with applications to color filtering *IEEE Trans. Image Process.* **5** 1582–6
[18] Tschumperlé D and Deriche R 2005 Vector-valued image regularization with pdes: a common framework for different applications *IEEE Trans. Pattern Anal. Mach. Intell.* **27** 506–17
[19] Bresson X and Chan T F 2008 Fast dual minimization of the vectorial total variation norm and applications to color image processing *Inverse Problems Imaging* **2** 455–84
[20] Goldluecke B, Strekalovskiy E and Cremers D 2012 The natural vectorial total variation which arises from geometric measure theory *SIAM J. Imaging Sci.* **5** 537–63
[21] Arridge S R and Simmons A 1997 Multi-spectral probabilistic diffusion using bayesian classification *Scale-Space Theories in Computer Vision* ed B M ter Haar Romeny, L Florack, J J Koenderink and M A Viergever (Berlin: Springer) pp 224–35
[22] Bilgic B, Goyal V K and Adalsteinsson E 2011 Multi-contrast reconstruction with bayesian compressed sensing *Magn. Reson. Med.* **66** 1601–15
[23] Leahy R M and Yan X 1991 Incorporation of anatomical MR data for improved functional imaging with PET *Information Processing in Medical Imaging* (Berlin: Springer) pp 105–20
[24] Bowsher J E, Johnson V E, Turkington T G, Jaszczak R J, Floyd C E and Coleman R E 1996 Bayesian reconstruction and use of anatomical a priori information for emission tomography *IEEE Trans. Med. Imaging* **15** 673–86
[25] Bowsher J E *et al* 2004 Utilizing MRI information to estimate F18-FDG distributions in rat flank tumors In *IEEE Nucl. Sci. Symp. Med. Imaging Conf.* 2488–92
[26] Chan C, Fulton R, Feng D D and Meikle S 2009 Regularized image reconstruction with an anatomically adaptive prior for positron emission tomography *Phys. Med. Biol.* **54** 7379–400

[27] Kazantsev D, Arridge S R, Pedemonte S, Bousse A, Erlandsson K, Hutton B F and Ourselin S 2012 An anatomically driven anisotropic diffusion filtering method for 3D spect reconstruction *Phys. Med. Biol.* **57** 3793–810

[28] Pedemonte S, Cardoso M J, Arridge S R, Hutton B F and Ourselin S 2012 Steady-state model of the radio-pharmaceutical uptake for MR-PET In *Med. Image Comput. Comput. Assist. Interv.* 289–97

[29] Somayajula S, Asma E and Leahy R M 2005 PET image reconstruction using anatomical information through mutual information based priors In *IEEE Nucl. Sci. Symp. Med. Imag. Conf.* 2722–6

[30] Nuyts J 2007 The use of mutual information and joint entropy for anatomical priors in emission tomography In *IEEE Nucl. Sci. Symp. Med. Imag. Conf.* 4149–54

[31] Tang J and Rahmim A 2009 Bayesian PET image reconstruction incorporating anato-functional joint entropy *Phys. Med. Biol.* **54** 7063–75

[32] Somayajula S, Panagiotou C, Rangarajan A, Li Q, Arridge S R and Leahy R M 2011 PET image reconstruction using information theoretic anatomical priors *IEEE Trans. Med. Imaging* **30** 537–49

[33] Ollinger J M and Fessler J A 1997 Positron-emisson tomography *IEEE Signal Process. Mag.* doi:10.1109/79.560323

[34] Bailey D L, Townsend D W, Valk P E and Maisey M N 2005 *Positron Emission Tomography-Basic Sciences* (London: Springer)

[35] Natterer F and Wübbeling F 2001 Mathematical methods in image reconstruction *SIAM Monographs on Mathematical Modeling and Computation.* (Philadelphia, PA: SIAM)

[36] Natterer F 2001 The mathematics of computerized tomography *Classics in Applied Mathematics.* (Philadelphia, PA: SIAM)

[37] Bertero M, Lantéri H and Zanni L 2008 Iterative image reconstruction: a point of view *Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy (IMRT)* ed Y Censor, M Jiang and A K Louis *CRM Series* vol 7, pp 37–63

[38] Liang Z and Lauterbur P C 2000 Principles of magnetic resonance imaging: a signal processing perspective *IEEE Press Series in Biomedical Engineering.* (Bellingham, WA: SPIE)

[39] McRobbie D W, Moore E A, Graves M J and Prince M R 2006 *MRI—From Picture to Proton* (Cambridge: Cambridge University Press)

[40] Gudbjartsson H and Patz S 1995 The rician distribution of noisy MRI data *Magn. Reson. Med.* **43** 910–4

[41] Macovski A 1996 Noise in MRI *Magn. Reson. Med.* **36** 494–7

[42] Dietrich O, Raya J G, Reeder S B, Reiser M F and Schoenberg S O 2007 Measurement of signal-to-noise ratios in MR images: influence of multichannel coils, parallel imaging, and reconstruction filters *J. Magn. Reson. Imaging* **26** 375–85

[43] Candès E J, Romberg J K and Tao T 2006 Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information *IEEE Trans. Inf. Theory* **52** 489–509

[44] Candès E J, Romberg J K and Tao T 2006 Stable signal recovery from incomplete and inaccurate measurements *Commun. Pure. Appl. Math.* **LIX** 1207–23

[45] Donoho D L 2006 Compressed sensing *IEEE Trans. Inf. Theory* **52** 1289–306

[46] Eldar Y C and Kutyniok G 2012 *Compressed Sensing: Theory and Applications* (Cambridge: Cambridge University Press)

[47] Lustig M, Donoho D L and Pauly J M 2007 Sparse MRI: the application of compressed sensing for rapid mr imaging *Magn. Reson. Med.* **58** 1182–95

[48] Ravishankar S and Bresler Y 2011 MR image reconstruction from highly undersampled k-space data by dictionary learning *IEEE Trans. Med. Imaging* **30** 1028–41

[49] Bishop C M 2006 *Pattern Recognition and Machine Learning* (Springer: Information Science and Statistics)

[50] Barber D 2012 *Bayesian Reasoning and Machine Learning* (Cambridge: Cambridge University Press)

[51] Pluim J P W, Maintz J B A and Viergever M A 2000 Image registration by maximization of combined mutual information and gradient information *IEEE Trans. Med. Imaging* **19** 809–14

[52] Haber E and Modersitzki J 2006 In medical image computing and computer-assisted intervention (*Intensity Gradient Based Registration and Fusion of Multi-Modal Images* 46) (Berlin: Springer) pp 726–33

[53] Tzimiropoulos G, Zafeiriou S and Pantic M 2011 Robust and efficient parametric face alignment In *IEEE Int. Conf. Comp. Vis.* 1847–54

[54] Blomgren P and Chan T F 1998 Color TV: total variation methods for restoration of vector-valued images *IEEE Trans. Image Process.* **7** 304–9

[55] Eldar Y C, Kuppinger P and Bolcskei H 2010 Block-sparse signals: uncertainty relations and efficient recovery *IEEE Trans. Signal Process.* **58** 3042–54

[56] van den Berg E and Friedlander M P 2010 Theoretical and empirical results for recovery from multiple measurements *IEEE Trans. Inf. Theory* **56** 2516–27

[57] Duarte M F and Eldar Y C 2011 Structured compressed sensing: from theory to applications *IEEE Trans. Signal Process.* **59** 4053–85

[58] Sprechmann P, Ramírez I, Sapiro G and Eldar Y C 2011 C-hilasso: a collaborative hierarchical sparse modeling framework *IEEE Trans. Signal Process.* **59** 4183–98

[59] Alvarez L, Lions P and Morel J 1992 Image selective smoothing and edge detection by nonlinear diffusion II *SIAM J. Numer.Anal.* **29** 845–66

[60] Weickert J 1998 *Anisotropic Diffusion in Image Processing* (Stuttgart: Teubner)

[61] Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica D: Nonlinear Phenomena* **60** 259–68

[62] Grasmair M and Lenzen F 2010 Anisotropic total variation filtering *Appl. Math. Optim.* **62** 323–39

[63] Byrd R H, Lu P, Nocedal J and Zhu C 1995 A limited memory algorithm for bound constrained optimization *SIAM J. Sci. Comput.* **16** 1190–208

[64] Richard H. Byrd, Peihuang Lu, Jose Luis Morales, Jorge Nocedal, and Ciyou Zhu 2014 L-BFGS-B: Software for Large-scale Bound-constrained Optimization (www.ece.northwestern.edu/~nocedal/lbfgs.html)

[65] Stephen Becker 2014 L-BFGS-B: Mex wrapper (www.mathworks.co.uk/matlabcentral/fileexchange/35104-lbfgsb-l-bfgs-b-mex-wrapper)

[66] Matthias Joachim Ehrhardt and Simon R Arridge 2014 Software for 'Joint reconstruction of PET-MRI by exploiting structural similarity' (www.cs.ucl.ac.uk/staff/ehrhardt/software.html)

[67] Pedemonte S, Bousse A, Erlandsson K, Modat M, Arridge S R, Hutton B F and Ourselin S 2010 GPU accelerated rotation-based emission tomography reconstruction In *IEEE Nucl. Sci. Symp. Med. Imaging Conf. (NSS/MIC)* pp 2657–61

[68] Jin B, Lorenz D A and Schiffler S 2009 Elastic-net regularization: error estimates and active set methods *Inverse Problems* **25** 115022

[69] Lu S and Pereverzev S V 2011 Multi-parameter regularization and its numerical realization *Numer. Math.* **118** 1–31

[70] Wang W, Lu S, Mao H and Cheng J 2013 Multi-Parameter Tikhonov regularization with the l0 sparsity constraint *Inverse Problems* **29** 065018